1. Introduction:

Les modèles de Markov cachés (en anglais Hidden Markov Model ou HMM) ont été introduits par BAUM et ses collaborateurs dans les années 1960-1970. Ce modèle est fortement apparent aux automates probabilistes. Un automate probabiliste est défini par une structure composée d'états et de transitions, et par un ensemble de distribution de probabilité sur la transition.

A chaque transition est associé un symbole d'un alphabet fini, ce symbole est généré à chaque fois que la transition est empruntée. Un HMM se définit également par une structure composée d'états et de transitions et par un ensemble de distribution de probabilité sur les transitions. La différence essentielle avec les PAS (processus aléatoire stochastique) est que la génération de symboles s'effectue sur les états, et non sur les transitions. De plus, on associé à chaque état non pas un symbole de l'alphabet, une illustration de la parenté entre HMM et PAS peut-être trouvée dans la propriété que n'importe quel HMM peut être simulé par un processus aléatoire du même nombre d'états. Notons que la propriété inverse n'est pas vraie.

Les HMMs sont utilisés pour modéliser des séquences d'observations. Ces observations peuvent être de nature discrète (par exemple les caractères d'un alphabet fini) ou continue (la fréquence d'un signal, une température…). Le premier domaine auquel les HMMs ont été appliqué est le traitement de parole au début des années 1970. [10]

2. Chaine de Markov cachée et l'extension vers le HMM :

Considérons un système qui peut se décrire comme, à n'importe quel temps, se trouve dans les N états distincts S_1, S_2, ……….$S_{N.}$

Le système voit le changement des états par t=1,2,…, et nous noterons l'état actuel à l'instant t par $q_{t.}$ Pour le cas d'une chaine de Markov discrète et de premier ordre, la description probabiliste est entre l'état courant et l'état précédent :

$$P\left[q_t = S_j \backslash q_{t-1} = S_i, q_{t-2} = S_k \dots\dots\dots\dots.\right] = P\left[q_t = S_j \backslash q_{t-1} = S_i\right] \dots\dots\dots\dots (3.1)$$

Dans le cas d'une probabilité stationnaire nous noterons :

$$a_{ij} = P\left[q_t = S_j \backslash q_{t-1} = S_i\right], 1 \leq i, j \leq N \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots.. (3.2)$$
$$a_{ij} \geq 0 \text{ et } \sum_{j=1}^{N} a_{ij} = 1 .$$

Ce processus stochastique est appelé modèle de Markov observable VMM puisque chaque état est lié correspond à une observation dans chaque instant. [10]

Le modèle a une extension vers le cas ou l'observation est une fonction probabiliste de l'état, comme résultat le modèle obtenu se nomme modèle de Markov caché HMM, qui est un double processus stochastique qui n'est pas observé (caché) mais peut être observé à l'aide d'un processus stochastique qui produit la séquence d'état.

3. Notations d'un HMM :

Le modèle λ définit par les paramètres suivants :
$\lambda = (A, B, \pi)$,

D'où :

A:matrice de transition.
B:matrice d'observation.
Π:matrice initial.

4. Description des éléments d'un HMM :

Le HMM est caractérisé par les éléments suivants :

- N, le nombre des états cachés d'un modèle :

On désigne l'ensemble des états par $S = \{S_1, S_2, \ldots\ldots S_N\}$.

L'état où se trouve le HMM à l'instant t est noté q_t $(q_t \in S)$.

- M, le nombre des symboles d'observation dans chaque état :

On désigne l'ensemble d'observation par $V = \{V_1, V_2, \ldots\ldots V_M\}$,

Un élément O_T de V désigne un symbole observé à l'instant t.

- Une matrice A de probabilités de transition entre les états de la chaine :

$\forall$ i,j $\in$ [1 . . . n] , $\forall$t $\in$ [1. . . T] :

$$a_{ij} = A(i,j) = P(q_{t+1} = S_j \mid q_t = S_i)$$

$$\forall\, i,j \in [1 \ldots n] \;, \forall t \in [1 \ldots T]$$

Avec : $\mathbf{a_{ij} \geq 0 \quad \forall i, j}$ et $\sum_{j=1}^{n} \mathbf{a_{ij} = 1}$.

- Une matrice B de probabilités d'observation :

$b_j(K)$ est la probabilité d'observer le symbole V_k quand le modèle se trouve dans l'état j soit :

$$b_j(K) = P\left(O_t = v_k \middle| q_t = S_j\right) \quad 1 \leq j \leq n, 1 \leq k \leq M \ldots\ldots\ldots\ldots\ldots\ldots\ldots (3.3)$$

Avec : $b_j(K) \geq 0 \quad \forall\, j, k$ et $\sum_{k=1}^{M} b_j(k) = 1$.

- Un vecteur π de densités de probabilité initiale :

$\pi = \{\pi_i\}_i = 1,2, \ldots\ldots n$, $\pi_i$représente la probabilité que l'état de départ du modèle soit l'état i.

Soit :

$$\pi_i = P(q_1 = s_i) \quad 1 \leq i \leq n \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (3.4)$$

Avec : $\pi_i = \geq 0 \ \forall i$ et $, \sum_{i=1}^{n} \pi_i = 1$.

5. La partie cachée et la partie observable :

Un HMM représente une séquence par deux suites de variable aléatoires l'une cachée et l'autre observable.

- La suite cachée correspond à la suite des états $q_1, q_2, \ldots, q_T$, notée :

Q (1 : T), où les q_i prennent leur valeur parmi l'ensemble des n états du modèle $\{S_1, S_2, \ldots, S_n\}$.

- La suite observable correspond à la séquence d'observations $O_1, O_2, \ldots, O_T$, notée O(1 : T), ou les O_i sont aussi fonction du temps et réalisent parmi un ensemble de M symboles observables $\{v_1, v_2, \ldots, v_M\}$.

On cherchera à construire des HMM pour lesquels les suites observables sont les exemples que l'on cherche à modéliser.

Les séquences observées sont définies comme des phrases sur l'alphabet $\{V_1, V_2, \ldots, V_M\}$

6. Un HMM engendre une séquence:

Un HMM peut être vu comme un processus permettant d'engendrer une séquence ; inversement, on peut considérer une séquence comme une suit d'observations sur un HMM. [8]

- Choix de l'état de départ :

t= 1, choisir l'état de départ $q_1 = S_i \ \ \pi_i$;

- Observation dans l'état sélectionné :

Choisir l'observation $O_t = V_k$ avec $b_i(k)$;

- Sélection de l'état suivant :

Transiter à l'état suivant $q_{t+1} = S_j$ à l'aide de a_{ij} ;

- Changement effectif d'état :

t= t+1 ;

Si $t < T$ alors retourner à l'étape 2 sinon STOP fsi

Exemple d'un HMM :

3 pièces truquées : HMM à 3 états.

N = 3 états, $S = \{S_1, S_2, S_3\}$.

M = 2 symboles, V= {p, f}.

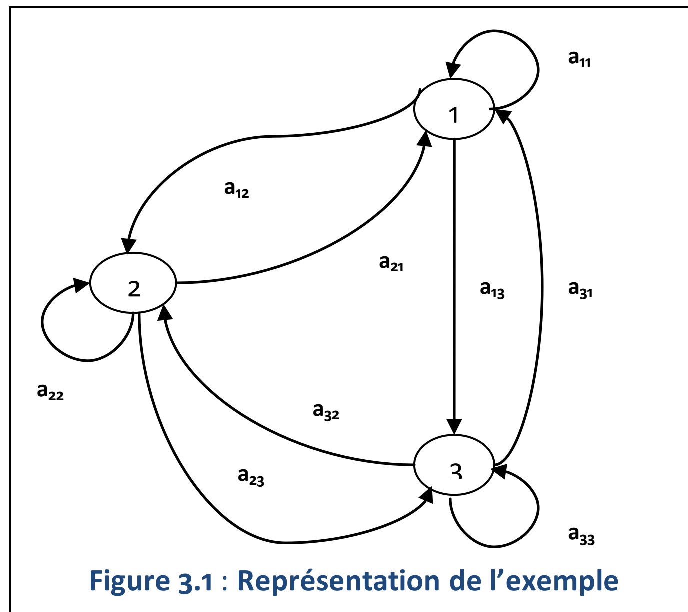


Figure 3.1 : Représentation de l'exemple

$$A = \begin{pmatrix} a_{11} & a_{21} & a_{31} \\ a_{21} & a_{22} & a_{32} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \qquad B = \begin{pmatrix} P_1 & 1 - P_1 \\ P_2 & 1 - P_2 \\ P_3 & 1 - P_3 \end{pmatrix}$$

7. Les trois problèmes de base d'un HMM :

IL ya trois problèmes à résoudre pour que la théorie des HMM_s puisse donner naissance à des algorithmes :

L'évaluation de la probabilité de l'observation d'une séquence : [10] [13]

Etant donnée la suite d'observations O et un HMM λ,

Comment évaluer la probabilité d'observation P (O \λ)?

La recherche du chemin le plus probable : [11]

Ou estimation de la partie cachée, ou encore décision. Soit la suite d'observation O et un modèle λ, comment trouver une suite d'états $Q = q_1,q_2, . . . q_T$ qui soit optimale selon un certain critère?

L'apprentissage : [10]

Comment ajuster les paramètres du modèle λ pour maximiser P(O\λ), à partir de séquences d'apprentissage dont on sait qu'elles ont été émises par ce modèle?

7.1 L'évaluation de la probabilité de l'observation d'une séquence : [10]

7.1.1 Evaluation directe de la probabilité d'observation :

La probabilité de la suite d'observations O, étant donnée le modèle λ est égale à la somme sur tous les chemins d'états possibles Q des probabilités conjointes de O et de Q:

$$P(O\backslash\lambda) = \sum_{Q} P(O,Q\backslash\lambda) = \sum_{Q} P(O\backslash Q,\lambda)P(Q\backslash\lambda) \dots\dots\dots\dots\dots\dots (3.5)$$

Or, on a les relations :

$$P(Q\backslash\lambda) = \pi_{q1}a_{q1q2}a_{q2q3}\dots\dots\dots . a_{q_{T-1}q_T}$$

$$P\,(O\backslash Q,\lambda) = b_{q1}(O_1)b_{q2}(O_2)\dots b_{qT}(O_T).$$

D'où :

$$P(Q\backslash\lambda) = \sum_{q1,q2\dots\dots qT} \pi_{q1}b_{q1}(O_1)a_{q1q2}b_{q2}(O_2)\dots\dots\dots a_{qT-1qT}b_{qT}b_{qT}(O_T)\dots\dots (3.6)$$

Cette formule directe nécessite $n^T - 1$ additions et $(2T - 1)n^T$ multiplications, soit $2Tn^T$ opérations.

7.1.2 Evaluation avec les fonctions de Forward-Backward :

On considère que l'observation peu se faire en deux temps:

d'abord, émission du début de l'observation O(1 : t) en aboutissant à l'état q_i au temps t, puis, émission de la fin de l'observation O(t + 1 : T) sachant que l'on part de q_i au temps t.

Dans ce cas, l'évaluation de l'observation est égale à :

$$P(O\backslash\lambda) = \sum_{q_i} \alpha(t,q_i)\beta(t,q_i) \dots\dots\dots\dots\dots\dots (3.7)$$

Ou α(t, qi) est la probabilité d'émettre de début O(1 : t) et d'aboutir à q_i à l'instant t, et $\beta(t,q_i)$est la probabilité d'émettre la fin O(t+1:T) sachant que l'on part de q_i à l'instant t. Forward-Backward: α pour t croissants, β pour t décroissant.

Le calcul de α :

$$\alpha_t(i) = P(O_1\ O_2\dots . O_t\ , q_t = S_i\backslash\lambda) \dots\dots\dots\dots\dots\dots (3.8)$$

$\alpha_t(i)$ est calculée de manière récursive comme suit :

- Initialisation :

$$\alpha_1(i) = \pi_i b_i(O_1)\ , 1 \le i \le n\ .$$

- Récurrence :

$$\alpha_{t+1}(j) = [\sum_{i=1}^{n} \alpha_t(i)a_{ij}]b_j(O_{t+1})\ .$$

- Terminaison :

$$P(O\backslash\lambda) = \sum_{i=1}^{n} \alpha_T(i) \ .$$

Ce calcul nécessite $n + n(n + 1)(T - 1)$ multiplications et $(n - 1)n(T - 1)$ additions, soit une complexité en $O(n2T)$.

Le calcul de β :

$$\beta_t(i) = P(O_{t+1}O_{t+2} \dots O_T \setminus q_T = s_i, \lambda) \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots (3.9)$$

On déduit $\beta_t$de β_{t+1} par :

- Initialisation :

$$\beta_T(i) = 1 , \quad 1 \leq i \leq n .$$

- Récurrence :

$$\beta_T(i) = \sum_{j=1}^{n} a_{ij}\, b_j(O_{t+1})\beta_{t+1}(j) . \quad 1 \leq t \leq T - 1 , 1 \leq i \leq n$$

$$P(O\backslash\lambda) = \sum_{qi} \alpha_t(i)\beta_t(i) \ .$$

Cependant, on utilise le plus souvent les valeurs obtenues pour deux cas particulières $(t = 0)$ ou $(t = T)$, ce qui donne :

$$P(O\backslash\lambda) = \sum_{qi} \alpha_T(i) = \sum_{qi} \pi_i \beta_0(i) \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots (3.10)$$

Exemple :

Reprenons le modèle $\lambda = (A , B , \pi)$ comprenant 3 états 1,2,3 et permet d'observer deux symboles a et b.

$S = \{S_1, S_2, S_3\}$.

$V = \{a, b\}$, calcul de α et $P(O\backslash\lambda)$, pour la suite.

$$A = \begin{pmatrix} 0,2 & 0,4 & 0,4 \\ 0 & 0,6 & 0,4 \\ 0 & 0 & 0 \end{pmatrix} \quad ,B = \begin{pmatrix} 0,1 & 0,5 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} \quad , \pi = \begin{pmatrix} 0,3 \\ 0,2 \\ 0,5 \end{pmatrix}$$

Suite d'observation : aab.

On a : $\alpha_1(i) = \pi(Oi)$ d'où :

$\alpha_1(1) = \pi_1\, b_1(a) = 0{,}3.0{,}5 = 0{,}15$.

$\alpha_1(2) = \pi_2 b_2(a) = 0{,}2.1 = 0.2$.

$\alpha_1(3) = \pi_3\, b_3(a) = 0{,}5$.

$\alpha_2(1) = [\sum_{i=1}^{3} \propto_1 (i)a_{i1}]\, b_1(a) = [\propto_1 (i)a_{11} + \propto_1 (2)a_{21} + \propto_1 (3)a_{31}]b_1(a)$
$= [0{,}15.0{,}2 + 0{,}2.0 + 0.1].0{,}5$

$\alpha_2(1) = 0{,}15.$

$\alpha_2(2) = [\sum_{i=1}^{3} \propto_i (i)\, a_{i2}]\, b_2(a) = [\propto_1 (1)a_{12} + \propto_1 (2)a_{22} + \propto_1 (3)a_{32}]b_2(a)\, b_2(a)$

$= [0{,}15.0{,}4 + 0{,}2.0{,}6].1$

$\alpha_2(2) = 0{,}18.$

$\alpha_2(3) = [\sum_{i=1}^{3} \propto_1 (3)a_{i3}]b_3(a) = [\propto_1 (1)a_{13} + \propto_1 (2)a_{23} + \propto_1 (3)a_{33}]b_3(a)$

$= [0{,}15.0{,}4 + 0{,}2.0{,}4].0$

$\alpha_2(3) = 0.$

$\alpha_3(1) = [\sum_{i=1}^{3} \propto_2 (i)a_{i1}]b_1(b) = [\propto_2 (1)a_{11} + \propto_2 (2)a_{21} + \propto_2 (3)a_{31}]b_1(b)$

$= [0{,}15.0{,}2 + 018.0].0{,}5$

$\propto_3 (1) = 0{,}015$.

$\alpha_3(2) = [\sum_{i=1}^{3} \propto_2 (i)a_{i3}]b_3(b) = [\propto_2 (1)a_{12} + \propto_2 (2)a_{22} + \propto_2 (3)a_{32}]b_2(b)$

$= [0{,}15.0{,}4 + 0{,}18.0{,}6].0$

$\alpha_3(2) = 0.$

$$\propto_3 (3) = \sum_{i=1}^{3} \propto_2 (i)a_{i3}b_3(b) = [\propto_2 (1)a_{13} + \propto_2 (2)a_{23} + \propto_2 (3)a_{33}]b_3(b)$$

$= [0{,}15.0{,}4 + 0{,}18.0{,}4].1$

$\alpha_3(2) = 0.132.$

$$P(O\backslash\lambda) = \sum_{i=1}^{3} \propto_3 (i) = \propto_3 (1) + \propto_3 (2) + \propto_3 (3)$$

$= 0{,}015 + 0{,}132$

$P(O\backslash\lambda) = 0{,}147.$

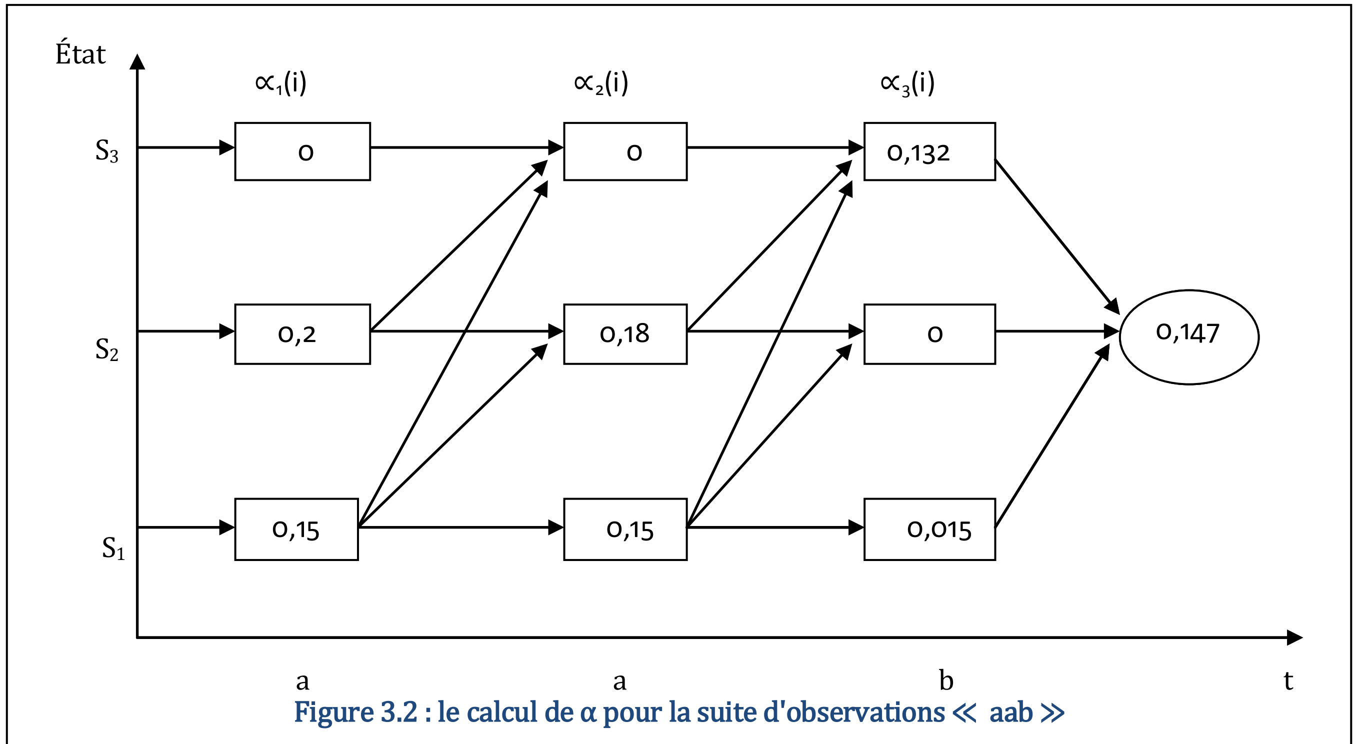


Figure 3.2 : le calcul de α pour la suite d'observations ≪ aab ≫

7.2 Calcul de chemin optimal: [11]

Il s'agit de trouver dans le modèle la suite d'états qui maximisent :

$$P(Q, O, \lambda) \Leftrightarrow P(Q, O \backslash \lambda).$$

Pour trouver le meilleur chemin $Q = (q_1, q_2 \dots \dots \dots, q_T)$ pour une suite d'observation

O= $(O_1, O_2, \dots \dots \dots, O_T)$, on définit $\delta_t(i)$ qui est la probabilité du meilleur chemin amenant à l'état S_i à l'instant t, en tenant guidé par les t premiers observations:

$$\delta_t(i) = \max_{q1,q2,.\,.\,.\,.\,.\,.\,.\,qt-1} P(q_1, q_2, \dots \dots \dots q_t = S_i\ , O_1, O_2, \dots \dots \dots\ O_t/\lambda) \dots \dots (3.11)$$

Par récurrence, on calcul $\delta_{t+1}(j) = \max[\delta_i(i)a_{ij}]\, b_j(O_{t+1})$.

On garde la trace de la suite d'états qui donne le meilleur chemin amenant à l'état S_i à t dans un tableau ψ.

7.2.1 Algorithme de Viterbi :

- Initialisation:

$\delta_1(i) = \pi_i b_i(O_1) \quad 1 \leq i \leq n$;

$\psi_t(i) = 0$;

- Induction:

$\delta_t(i) = \max[\delta_{t-1}(i)a_{ij}]\, b_j(O_t), \quad 1 \leq j \leq n, 1 \leq i \leq n$;

$\psi_t(i) = \arg\max[\delta_{t-1}(i)a_{ij}]$, $1 \leq j \leq n, 1 \leq i \leq n$;

- Terminaison:

$P^* = \max[\delta_T(i)]$; $1 \leq i \leq n$;

$q_T^* = \arg\max[\delta_T(i)]$, $1 \leq i \leq n$;

- Chemin obtenu : ("backtracking")

$q_t^* = \psi_{t+1}(q_{t+1}^*)$, $1 \leq t \leq T-1$;

Exemple:

En reprenant l'exemple précédent, on peut calculer les quantités δ, Ψ, q* puis les autres quantités aussi:

- Initialisation:

$\delta_1(1) = \pi_1 b_1(a) = 0{,}3.0{,}5 = 0{,}15.$ $\psi_1(1) = 0;$

$\delta_1(2) = \pi_2 b_2(a) = 0{,}2.1 = 0{,}2.$ $\psi_1(2) = 0;$

$\delta_1(3) = \pi_3 b_3(a) = 0.$ $\psi_1(3) = 0;$

- induction:

$\delta_2(1) = \max[\delta_1(i)a_{i1}]\, b_1(a).$

$= \max[\delta_1(1)a_{11}; \delta_1(2)a_{21}; \delta_1(3)a_{31}]\, b_1(a).$

$= \max[\underline{0{,}15.0{,}2}\,; 0{,}2.0\,; 0].0{,}5$.

$\delta_2(1) = 0{,}015$; $\psi_1(1) = 1$

$\delta_2(2) = \max[\delta_1(i)a_{i2}]\, b_2(a).$

$= \max[\delta_1(1)a_{12}; \delta_1(2)a_{22}; \delta_1(3)a_{32}]\, b_2(a).$

$= \max\, [0{,}15.0{,}4\,; \underline{0{,}2.0{,}6}\,; 0]\, .1$.

$\delta_2(2) = 0{,}12$; $\psi_2(2) = 2$.

$\delta_2(3) = \max[\delta_1(i)a_{i3}]\, b_3(a).$

$= \max[\delta_1(1)a_{13}; \delta_1(2)a_{23}; \delta_1(3)a_{33}]\, b_3(a).$

$= \max\, [0{,}15.0{,}2\,; \underline{0{,}2.0{,}4}\,; 0]\, .0$.

$\delta_2(3) = 0$; $\psi_2(3) = 2$

$\delta_3(1) = \max[\delta_2(i)a_{i1}]\, b_1(b).$

$= \max[\delta_2(1)a_{11}; \delta_2(2)a_{21}; \delta_2(3)a_{31}]\, b_1(b).$

$= \max\ [\underline{0{,}015.0{,}2}\ ;0{,}12.0\ ;0]\ .0{,}5\ .$

$\delta_3(1) = 0{,}0015 \quad ;\ \psi_3(1) = 1\ .$

$\delta_3(2) = \max[\delta_2(i)a_{i2}]\, b_2(b).$

$= \max[\delta_2(1)a_{12}\ ; \delta_2(2)a_{22}\ ; \delta_2(3)a_{32}]\, b_2(b).$

$= \max[0{,}015.0{,}4\ \ ;0{,}12.0{,}6;\ 0]\ .0\ .$

$\delta_3(2) = 0 \quad ;\ \psi_3(2) = 2.$

$\delta_3(3) = \max[\delta_2(i)a_{i3}]\, b_3(b).$

$= \max[\delta_2(1)a_{13}; \delta_2(2)a_{23}; \delta_2(3)a_{33}]\, b_3(b).$

$= \max[0{,}015.0{,}4\ \ ;\underline{0{,}2.0{,}4};\ 0]\ .1\ .$

$\delta_3(3) = 0{,}08 \quad ;\ \psi_3(3) = 2\ .$

- Terminaison:

$\psi_1(1) = 0 \qquad \psi_2(1) = 1 \qquad \psi_3(1) = 1$

$\psi_1(2) = 0 \qquad \underline{\psi_2(2) = 2} \qquad \psi_3(2) = 2$

$\psi_1(3) = 0 \qquad \psi_2(3) = 2 \qquad \underline{\psi_3(3) = 2}$

$q_3^* = max\ [\delta_3(1); \delta_3(2); \delta_3(3)] = 3\ .$

- Chemin obtenu:

$$q_2^* = \psi(q_3^*) = \psi_3(3) = 2\ .$$

$$q_1^* = \psi(q_2^*) = \psi_2(2) = 2\ .$$

On déduit du tableau précédent que le meilleur chemin est : 2 2 3.

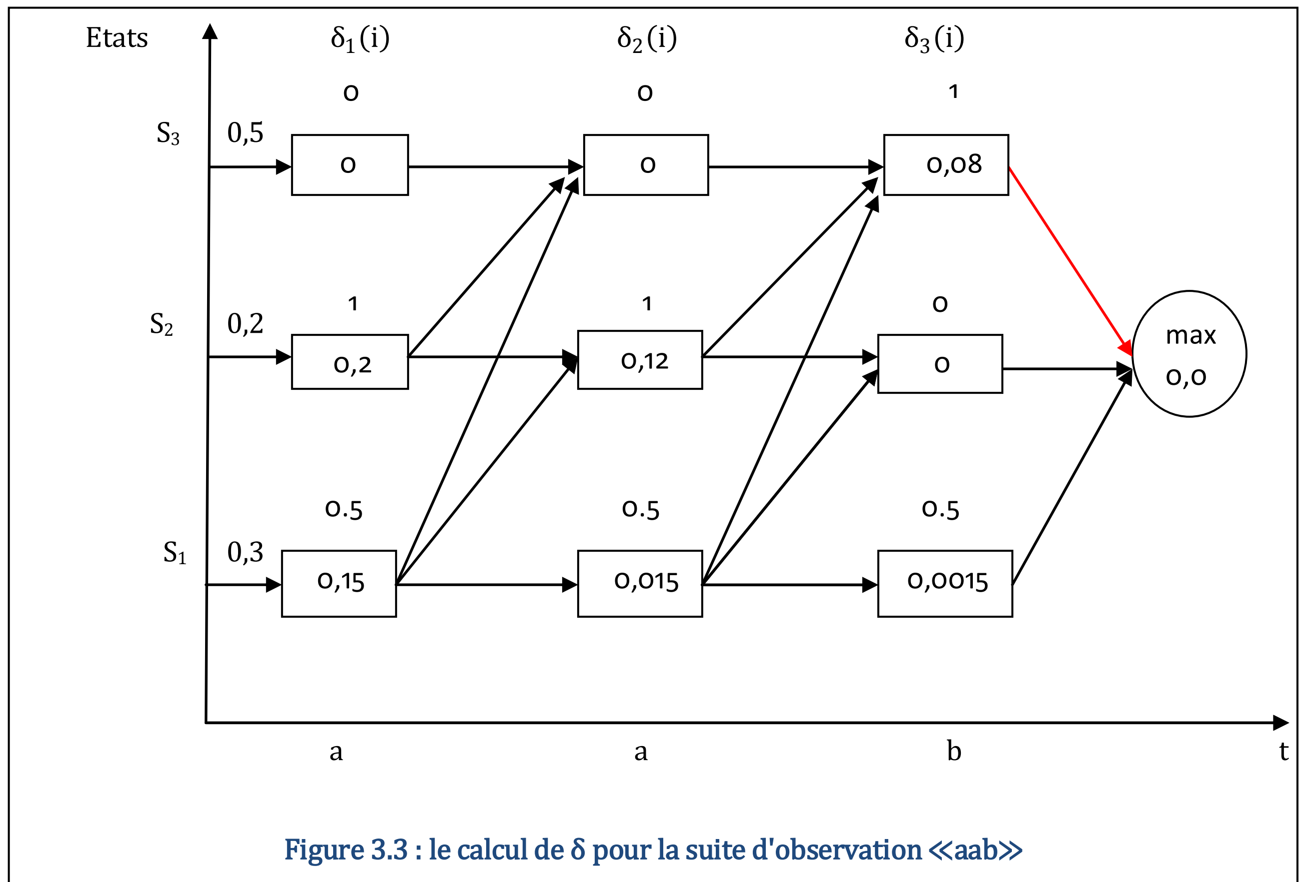


Figure 3.3 : le calcul de δ pour la suite d'observation ≪aab≫

7.3 Apprentissage: [10]

Le but de l'apprentissage est de déterminer les paramètres (A, B, π) qui maximisent la probabilité de la suite d'observation $P(O\backslash\lambda)$.

L'idée est d'utiliser des procédures de ré-estimation par punition récompense:

- choisir un ensemble initial de paramètres λ_0 ;
- calculer λ_1 à partir de λ_0;
- répéter ce processus jusqu'à un critère de fin.

Partant de λ_n ; λ_{n+1} doit vérifier :

$$P = (O\backslash\lambda_{n+1}) \geq P(O\backslash\lambda_n)$$

On cherche à définir une fonction F telle que $\lambda_{n+1} = F(\lambda_n)$.

L'approche la plus simple pour trouver une fonction F consiste à faire des statiques sur l'utilisation des transitions et des distributions. Ceci revient à calculer des fréquences d'utilisation à partir de l'ensemble d'apprentissage.

Si l'ensemble est important, ces fréquences fournissent une bonne approximation des probabilités à postériorités utilisables alors comme paramètres du modèle pour l'itération suivante.

La méthode d'apprentissage va donc consister à partir d'un modèle initial aléatoire à estimer ses paramètres comme indiqués ci-dessus, puis à recommencer cette estimation ("ré-estimation") jusqu'à obtenir une certaine convergence.

7.3.1 Les formules de ré-estimation:

On définit $\xi_t(i,j)$ comme la probabilité, étant donné une phrase O, et un HMM λ. Que se soit l'état S_i qui ait émis la lettre de rang t de O et l'état S_j qui ait émis celle de rang t+1, donc:

$$\xi_t(i,j) = P\left(q_t = S_i, q_{t+1} = S_j \backslash O, \lambda\right) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (3.12)$$

Où autrement, on peut écrire:

$$\xi_t(i,j) = \frac{P\left(q_t = S_i, q_{t+1} = S_j / O, \lambda\right)}{P(0/\lambda)}$$

À partir de ces fonctions Forward-Backward, on en déduit:

$$\xi_t(i,j) = \frac{\alpha_t(i)\alpha_{ij}\, b_j(O_{t+1})\beta_{t+1}(j)}{P(0/\lambda)}$$

On définit aussi la quantité $\gamma_t(i)$ comme la probabilité que la lettre du rang de la phase O soit émise par l'état S_j

$$\gamma_t(i) = P(q_t = S_i / O, \lambda)$$

Soit:

$$\gamma_t(i) = \sum_{j=1}^{n} P(q_t = s_i, q_{t+1} = S_j / O, \lambda) = \frac{\sum_{j=1}^{n} P(q_t = s_i, q_{t+1} = S_j, O, \lambda)}{P(0/\lambda)}$$

On a la relation:

$$\gamma_t(i) = \sum_{j=1}^{n} \xi_t(i,j) = \frac{\alpha_t(i)\beta_t(i)}{P(0/\lambda)} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (3.13)$$

Le nouveau modèle HMM se calcul à partir de l'ancien en ré-estimant π, A et B Par comptage sur la base d'apprentissage. On mesure les fréquences:

$$\pi_i = \text{nombre de fois que HMM s'est trouvé dans } S_i \text{à l'instant } t = 1.$$

$$a_{ij} = \frac{\text{nombre de fois ou la transition de Si à Sj à été utilisée}}{\text{nombre de transition èffectuées àpartir de Si}}$$

$$b_j(k) = \frac{\text{nombre de fois ou le HMM s'est trouvé dans l'étatSj en obsevant vk}}{\text{nombre de fois ou HMM s'esttrouvé dans l'état Sj}}$$

$$\pi_i = \frac{1}{N}\sum_{k=1}^{N}\gamma 1(i) \ldots\ldots\ldots\ldots\ldots\ldots (3.14)$$

$$a_{ij} = \frac{\sum_{k=1}^{N}\sum_{t=1}^{0k}\xi t(i,j)}{\sum_{k=1}^{N}\sum_{t=1}^{0k-1}\gamma t(i)} \ldots\ldots\ldots\ldots\ldots\ldots (3.15)$$

$$b_{ij}(1) = \frac{\sum_{k=1}^{N}\left(\sum_{t=1,0t=v1}\gamma t(j)\right)}{\sum_{k=1}^{N}\sum_{t=1}^{0k-1}\gamma t(j)} \ldots\ldots\ldots\ldots\ldots\ldots (3.16)$$

L'algorithme de Baum-Welch:

- fixer les valeurs initiales:

α_{ij}^{0} , $b_j^0(k)$, π_i^0. $1 \leq i , j \leq n$, $1 \leq k \leq n$.

- Calcul à l'aide des fonctions de Forward-Backward:

$\xi(i,j), \gamma_t(i)$, $1 \leq i,j \leq n$, $1 \leq t \leq T-1$.

Et λ en utilisant les formulas de ré-estimation.

- recommencer en 2 jusqu'à ce certain critère de convergence soit rempli.

Exemple:

Soit un HMM λ_0 définit par les paramètres suivants:

$$A = \begin{pmatrix} 0{,}45 & 0{,}35 & 0{,}2 \\ 0{,}1 & 0{,}5 & 0{,}4 \\ 0{,}15 & 0{,}25 & 0{,}6 \end{pmatrix} \qquad B = \begin{pmatrix} 1 & 0 \\ 0{,}5 & 0{,}5 \\ 0 & 1 \end{pmatrix} \qquad \pi = \begin{pmatrix} 0{,}5 \\ 0{,}3 \\ 0{,}2 \end{pmatrix}$$

On peut calculer sur l'alphabet à deux lettres $\{a, b\}$:

P (abbaa\\λ_0) = 0,0278.

Si on prend comme ensemble d'apprentissage cette seule phrase, l'application de l'algorithme de Baum-Welch doit augmenter sa probabilité de reconnaissance.

Après une ré-estimation, on trouve λ_1 :

$$A = \begin{pmatrix} 0{,}346 & 0,365 & 0{,}289 \\ 0{,}159 & 0{,}514 & 0{,}327 \\ 0{,}377 & 0{,}256 & 0{,}364 \end{pmatrix} \qquad B = \begin{pmatrix} 1 & 0 \\ 0{,}5 & 0{,}5 \\ 0 & 1 \end{pmatrix} \qquad \pi = \begin{pmatrix} 0{,}656 \\ 0{,}344 \\ 0 \end{pmatrix}$$

P (abbaa\\λ_1) = 0,0529.

Après 15 itérations:

$$A = \begin{pmatrix} 0 & 0 & 1 \\ 0{,}212 & 0{,}788 & 0 \\ 0 & 0{,}515 & 0{,}485 \end{pmatrix} \qquad B = \begin{pmatrix} 1 & 0 \\ 0{,}969 & 0{,}031 \\ 0 & 1 \end{pmatrix} \qquad \pi = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$$

P $(abbaa \backslash \lambda_{15}) = 0{,}2474$.

Après 150 itérations la convergence est réalisée:

$$A = \begin{pmatrix} 0 & 0 & 1 \\ 0{,}18 & 0{,}82 & 0 \\ 0 & 0{,}5 & 0{,}5 \end{pmatrix} \qquad B = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} \qquad \pi = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$$

P $(abbaa \backslash \lambda) = 0{,}25$.

8. Principaux types de HMM : [12]

Il existe deux principaux types de modèles de Markov cachés, le modèle ergodique et le modèle gauche-droit.

8.1 Le modèle ergodique :

Tout état est atteignable depuis tout autre état en un nombre fini de transitions.

Exemple pour N = 4:

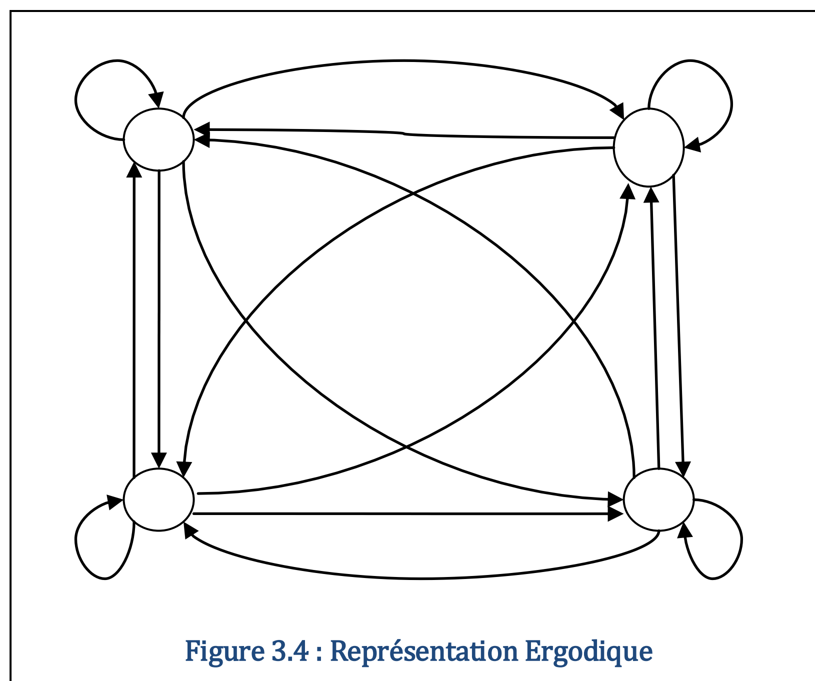


Figure 3.4 : Représentation Ergodique

8.2 Le modèle gauche-droit:

Si t augmente, alors les indices des états augmentent également.

Le modèle le plus connu est le modèle de Bakis:

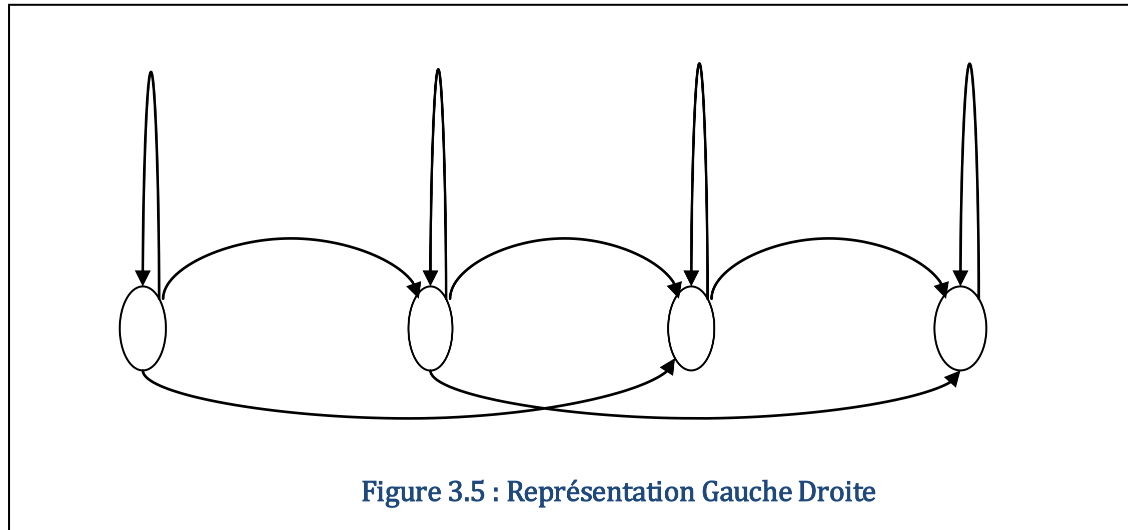


Figure 3.5 : Représentation Gauche Droite

Formellement, dans un modèle gauche-droit:

$a_{ij} = 0 \text{ si } j < i$

$\pi_i = 0 \text{ si } i \neq 1 , \pi_i = 0 \text{ si } i = 1 .$

De même on a souvent des contraintes supplémentaires comme:

$a_{ij} = 0 \text{ si } j > i + \Delta(\Delta = 2 \; dans \; les \; modéles \; de \; Bakis)$

9. Avantages et inconvénients des HMM: [12]

Parmi les inconvénients de ces modèles, on peut citer:

- Ignorance complète de la durée relative des événements dans le signal.
- Dégradation des performances si l'apprentissage n'est pas suffisant.
- Le choix a priori de la typologie des modèles(le nombre d'états, les transitions autorisées et les règles de transition) limite de la souplesse des modèles.
- Ce n'est qu'un modèle comportemental et non fonctionnel.
- Difficulté d'inclure des connaissances acquises.

Par contre, ils présentent un bon nombre d'avantages:

- Base mathématique solide pour comprendre le fonctionnement.
- Bon modèle de la variable de la forme.
- L'alignement temporel incorporé systématiquement.

- Prise en compte de l'ordre d'apparition dans une séquence de vecteurs.
- Reconnaissance réalisée par un simple calcul de probabilité cumulée. L'algorithme de VITERBI permet de la calculer efficacement, cet algorithme peut se formuler en terme de programmation dynamique (DTW).
- Décision globale sans obligation d'utiliser des seuils.
- Séparation franche entre données et algorithme.

10. Les principaux domaines d'applications: [12]

- Les transports et le trafic (dessertes par autobus, modèle de trafic routier, files d'attente, trafic dans les réseaux téléphonique hertziens ou télématiques, etc...).
- La théorie économique (modèles financiers, la théorie de la provision).
- Les sciences de l'environnement (pollution, cycles écologiques, génétique des population, etc....).
- La théorie de l'information, du filtrage et de la commande optimale (codage, estimation, détection, etc...).
- L'interprétation des signaux (la reconnaissance vocale, manuscrite et visuelle, l'identificateur du locuteur, etc...).

11. conclusion :

Dans ce chapitre nous avons défini d'une façon générale les modèles de Markov cachés et leurs applications dans plusieurs domaines.

Le progrès des cette technique envisage dans l'avenir de véritables systèmes de dialogue homme-machine dans le cadre de la reconnaissance automatique de la parole ou les recherches sont avancées. Et notons que cette approche probabilité (HMM) a une vaste utilisation dans plusieurs applications.

Le principe d'utilisation de ces méthodes est simple, il s'agit de déterminer, dans un modèle préalablement appris, la séquence d'états correspondant à une observation donnée.